DOUGLASS J. WILDE

OPTIMUM SEEKING METHODS

PRENTICE-HALL INTERNATIONAL SERIES
in the PHYSICAL and CHEMICAL ENGINEERING SCIENCES

Optimum Seeking Methods

DOUGLASS J. WILDE
Department of Chemical Engineering Stanford University

PRENTICE-HALL, INC.
Englewood Cliffs, N.J.

PRENTICE-HALL INTERNATIONAL, INC., *London*
PRENTICE-HALL OF AUSTRALIA, PTY., LTD., *Sydney*
PRENTICE-HALL OF CANADA, LTD., *Toronto*
PRENTICE-HALL FRANCE, S.A.R.L., *Paris*
PRENTICE-HALL OF INDIA (PRIVATE) LIMITED, *New Delhi*
PRENTICE-HALL OF JAPAN, INC., *Tokyo*
PRENTICE-HALL DE MEXICO S.A., *Mexico City*

Library of Congress Catalog Card Number 63-20039

Printed in the United States of America

63823 C

Preface

Although this book is based primarily on the work of statisticians, the author is a chemical engineer. This may suggest the interdisciplinary character of the work, which hopefully is broad enough in viewpoint to interest not only engineers and statisticians, but economists, applied mathematicians, operations analysts, and behavioral scientists as well.

Optimization—finding the best way to do things—is obviously of interest in the practical world of production, trade, and politics, where small changes in efficiency can spell the difference between success or disaster for any enterprise, be it neighborhood store, mammoth industrial complex, or governing political party. Today as always many important decisions are made simply by describing the system under study as precisely and quantitatively as possible, selecting some measure of system effectiveness, and then seeking the state of the system which gives the most desirable value of this criterion. Since description and understanding of systems is the traditional task of the engineers, economists, and other applied scientists for whom this book is written, it would be presumptious to discuss such a broad topic in this brief work. Moreover, choosing a measure of effectiveness is in most cases either completely obvious or so clouded by nonquantitative value judgments as to be extremely difficult, and so we do not deal with this delicate question either. Instead we concentrate entirely on the technical problems associated with the process of optimization itself.

Over a span of almost two centuries, the only mathematical methods known for handling optimization problems were the classical differential and variational calculus. With the rise of "operations research" since the Second World War, there has been renewed interest in optimization methods for dealing with problems not solvable by classical methods. Many of these techniques—linear and dynamic programming, for instance—have already been extensively described in technical books. However, a search for the optimum of a function more or less unknown to the observer necessarily involves experimentation, for the only way to gain information about such a function is by direct measurement. For this reason we have given the name *optimum seeking procedures* to the strategies guiding search for the optimum of any function about which full knowledge is not available. Such functions arise not only in situations where direct observations must be made on a physical or economic system, but also in theoretical studies where the mathematical form of the criterion of effectiveness is so complicated that it can only be evaluated directly on a high speed computer.

This book is based on material presented to senior engineering students whose mathematical training comprised only standard calculus courses, and anyone who can take a partial derivative should be also to follow our exposition. Indeed, the development depends very little on the calculus, the mathematics being rather simple, although probably a bit unfamiliar to engineers and economists. The book is intended not only as a text for undergraduate science students, but also as a reference book for practicing economists, engineers, and statisticians who may not have been able to keep up with the rapid development of experimental optimization techniques.

Problems in experimental optimization are good vehicles for learning about logic, multidimensional geometry, and elementary probability theory. As consequences of the technical treatment certain general decision principles are developed which may be useful even when there is no time for detailed and rigorous analysis. For example, study of experimental optimization problems involving a single independent variable gives insight into the important minimax concept as well as the somewhat startling technique of randomization. Close examination of multivariable problems unearths some rather disturbing facts about graphical reasoning and the paradoxes that can arise from failing to realize that even engineers and economists often must work with non-Euclidean space. Analysis of interactions between variables shows how one may blunder onto a false optimum—one which appears optimal but really isn't. Stochastic approximation theory illustrates just how experimental error can slow down a search. It also indicates how one might weight new observations with old ones to improve operations. As might be expected, changes should be more and more gradual as experience is gained, and the harmonic sequence $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$ turns out to be particularly significant in weighting successive observations to cancel out random errors.

The style of presentation has been greatly influenced by the ideas of G. Polya of Stanford University in his books *How to Solve it* and *The*

Mathematics of Plausible Inference. Thus we have tried to show how the originators of the techniques discussed might have discovered them. As Polya suggests, this approach, while requiring more words than a formal proof, helps develop creativity in the student. We have found that presenting the material this way has in fact stimulated students to original and valuable ideas, and it is a pleasure to acknowledge here the contributions of undergraduates Roger Ben Haïm at the Ecole Nationale Supérieure des Industries Chimiques, Nancy, France, and of Leaton T. Oliver and Rafael B. Cruz-Diaz at the University of Texas. While acknowledging the help of others, let us thank Professors J. Kiefer of Cornell University, Thomas E. Corrigan of Ohio State, Irving F. Miller of the Polytechnic Institute of Brooklyn, and William A. Graves of Colorado State, who have caught several mistakes and suggested better methods of presentation. Particular credit is due professors Robert J. Buehler, B. V. Shah, and Oscar Kemphorne of Iowa State University, who let us see their material on the method of parallel tangents before its publication. Industrial colleagues have also made worthy contributions: S. M. Johnson of the RAND Corporation, Robert Hooke, T. A. Jeeves, and Cy Wood of Westinghouse, Ed Blum of Pure Oil, Shahen Hovanessian and the group at T-R-W Computers, and our old friend Gene Motte of Union Oil.

We must also thank the colleagues who brought important material to our attention which might otherwise have been overlooked: Leroy Folks of Oklahoma State, Jim Carley of The University of Arizona, David Himmelblau of The University of Texas, Dale Rudd of The University of Wisconsin, and Alvin Harkins of the Monsanto Chemical Company. A citation for courage should be awarded Professors Robert Adler and Earl Gosė of the Case Institute of Technology, who were the first to use this book for part of a course on optimization theory. The author is also indebted to the institutions which supported him both morally and financially during the conception, classroom presentation, and writing of the book. They are, in order of appearance, The U.S. Education Commission in France, who supported a Fulbright lectureship at the Ecole Nationale Supérieure des Industries Chimiques in Nancy, France; the Israel Institute of Technology (Technion) in Haifa, Israel; The University of Texas, Austin; the National Science Foundation, who invited the author to present this material at the 1962 Process Dynamics and Optimization summer course for Engineering Professors held at the University of Colorado in Boulder; and Yale University. To the shade of the poet Henry Wadsworth Longfellow, who in his poem "Excelsior" also considered the problem of attaining the heights, the author offers his humble respects.

DOUGLASS J. WILDE

Contents

Search Problems 1

The shades of night were falling fast,
As through an Alpine village passed
A youth, who bore, 'mid snow and ice,
A banner with the strange device,
Excelsior!
—Longfellow

A scientist confronted by a system more or less unknown to him gains knowledge about it by making experiments. He fixes the variables under his control at settings of his choice, notes down the values of any factors he is unable to regulate, and then observes the results. In general, of course, there may be additional variables influencing the outcome which can be neither measured nor controlled; these are the factors behind the *experimental error*. Scientists are not the only people concerned with studying unknown systems, as Table 1-1 illustrates.

Although an experimental approach has always characterized the empirical sciences and may date all the way back to Eve's historic investigation of the apple in the Garden of Eden, there recently has been a renewal of interest in the basic nature of such investigations, often called "black-box" problems for reasons shown in the last line of Table 1-1. The next to last line, dealing with the mathematician's study of functions, gives the nomenclature and notation we shall use in discussing this problem.

In this book we are not particularly interested in finding out all there is to know about the system. We merely wish to determine what settings of the independent variables will yield the optimal (maximum or minimum, whichever is desired) value of whatever dependent variable is taken as a criterion of effectiveness. Let us then consider optimization problems in which the functional dependence of the efficiency criterion upon the adjustable variables is not known. Assume further that the total number n of experiments we can perform is limited. Any set of instructions for

Table 1-1

INVESTIGATION OF UNKNOWN SYSTEMS

Investigator	*System*	*Adjustable Factors*	*Uncontrollable but Measurable Factors*	*Unknown Factors*	*Results*
Farmer	Farmland	Fertilizer	Weather	Soil condition	Crop quality
Physician	Patient	Medicine	Pulse rate	Infection	Blood count
Engineer	Chemical reactor	Temperatures	Raw material composition	Activity of catalyst	Product yield
Detective	Suspect	Interrogation	Material evidence	Character of suspect	Testimony
Senator	Senate	Speech	Current events	Personalities of colleagues	Votes on bill
Sales manager	Market	Price of product	Competitors' prices	Public taste	Sales
Mathematician	Function $f(x, p, r)$	Independent variables x	Parameters p	Random variables r	Dependent variable $y=f(x, p, r)$
Anybody	"Black box"	Input knobs	Gauges	State of box	Output readings

placing the n experiments $x_1, x_2, \ldots, x_n$ will be called a *search plan,* and any investigation seeking the optimal value of an unknown function will be called a *search problem.* We want of course to find, from among all possible search plans, the one which looks for this optimum in an optimal manner. Thus we are not only trying to optimize the *function*; we are also optimizing the *optimization procedure.*

Search problems occur for several reasons. The theory describing a real system is rarely perfect, and many approximations are made during conception, design, and construction. Moreover, important parameters may change as time passes. And even with perfect theory and invariant parameters, errors of measurement often obscure the true relationship between output and input. While scientific theory is useful for finding nearly optimal conditions, the preliminary estimate often can be improved by experimenting directly with the system itself, be it a set of equations, a steel mill, or a truck farm.

1.01. Types of search problem

Search problems can be classified according to their number of independent variables and to whether or not experimental error is present. Except for the last chapter on stochastic approximation, we will deal with the error-free or deterministic case. When there is but one independent variable and no random error, elegant and powerful methods for finding an optimum are available; these will be described in Chap. 2. Unfortunately such procedures cannot be extended to problems with more than one independent variable, and many different techniques have been developed for handling the multivariable case. These will be described for the most part in Chaps. 4 and 5 after an introduction to multidimensional geometry in Chap. 3. The presence of experimental error slows down a search considerably and nullifies the essential advantage that univariable problems had over multivariable ones in the deterministic case. Thus in Chap. 6 all types of error-ridden situation are treated alike, regardless of the number of variables.

1.02. Roots and peaks

The problem of locating a peak is very much like that of finding a root. Many root-finding situations, where computationally convenient, have been formulated as minimization problems. Consider, for example, m simultaneous equations

$$\varphi_j(x_1, \ldots, x_k) = 0 \qquad (j = 1, 2, \ldots, m)$$

in the k unknowns x_1 through x_k. One can find solutions to this system of equations by minimizing the function Φ obtained by adding the squares of the left members φ_j. More properly, since the solutions may be complex, the function to be minimized should be

$$\Phi \equiv \sum_{j=1}^{k} \varphi_j \bar{\varphi}_j$$

where $\bar{\varphi}_j$ is the complex conjugate of φ_j. If $k = m$, then at the solution Φ will be at its minimum value zero. But if there are more equations than variables, as is usually the case in curve fitting problems, the minimum value of Φ may not be zero. Such a "solution" would represent a least-squares approximation.†

Just as root problems can be transformed into exercises in minimization, optimization problems can be solved by root-finding techniques. This is because the first derivatives of a continuous function must vanish at an extreme point. Hence by working with derivatives of the function to be optimized one can sometimes use well-known root-finding procedures to locate the optimum. The reader should keep in mind that the methods developed in this book, while specially suited to optimization problems, could perhaps be adapted to root location as well.

† A. D. Booth, "An Application of the Method of Steepest Descents to the Solution of Systems of Nonlinear Simultaneous Equations," *Quart. J. Mech. Appl. Math.*, 2, 4 (1949), p. 460.

1.03. Deterministic problems

Search problems having no unknown or random factors will be described as *deterministic*. Assuming that the investigator can correct for any uncontrollable but known factors affecting the system under study, we lose no generality in treating a simplified system having only adjustable variables on which depends some criterion y to be optimized.

In many systems the experimental error cannot be neglected. Yet there are interesting practical optimization problems that are wholly deterministic, as when the criterion of effectiveness is too involved to be optimized directly by such standard techniques as the differential calculus. Economic studies and engineering design problems for example, must usually be considered search problems because of the complexity of the mathematical model used—the equations, tables, graphs, or computer codes. When, as is often the case, the model involves no random elements, the search problem is entirely deterministic, even if the key figures are only approximations based on shaky assumptions, rules of thumb, and imperfect data. It would of course be unwise to expend very much searching effort on a model of questionable accuracy. But this does not affect our problem here, which is how to conduct the search effectively.

1.04. Stochastic problems

If experimental error cannot be neglected, the problem will be called *stochastic*, a word meaning that random factors are involved. Stochastic problems are naturally more difficult than deterministic ones, although not as hard as one might think at first. We shall see that it is possible to consider a stochastic problem as a deterministic one with noise, or experimental error, superimposed. In this way the problem of convergence of the deterministic part can be treated separately from that of the nullification of the noise.

The main effect of random error is to slow down the speed at which a search can be conducted and still be sure of eventually finding the optimum. Stochastic procedures, being very deliberate, should not be used in the absence of experimental error, for deterministic methods are much faster. This point has not been well understood in the past, and stochastic procedures have sometimes been applied to deterministic problems with disappointing results. One can avoid misapplication by remembering that stochastic methods must face a convergence problem that simply does not arise when there is never any danger of a mistake. Thus *speed* of convergence, which is the only consideration in deterministic problems, is only of secondary importance when there is noise.

1.05. Simultaneous and sequential procedures

Search plans fall naturally into two mutually exclusive classes which we shall call *simultaneous* and *sequential.* Plans specifying the location of every experiment before any results are known will be called *simultaneous,* while a plan permitting the experimenter to base future experiments on past outcomes will be called *sequential.*

Suppose, for illustration, that ten economists were available to analyze a capital investment program. Let us say that each economist, if given a tentative distribution of the capital, can, after about a week of calculation, estimate the return on the investment. If the group has only a week to prepare a recommendation, then a ten experiment simultaneous search plan is needed to pick the cases to be studied, for no one has the time to wait for the results of one case before choosing the conditions for another.

When this can be done, it is much more effective to use a *sequential* search plan. If the ten economists could team up and together analyze a case in half a day (assuming a five day week for economists), they could locate the optimum as effectively after analyzing only four cases sequentially as they could by analyzing ten simultaneously. And if instead of going home on Tuesday they continued analyzing cases all week, their ten sequential cases would be almost eighteen times as effective in locating the optimum as would the ten simultaneous cases—provided the sequential search plan were the optimal one. We shall see that the advantage of sequential plans over simultaneous plans increases exponentially with the number of experiments. The branch of statistics known as *experimental design* is usually concerned more with simultaneous than with sequential procedures, and since many good texts on this subject are available, we shall not spend very much time on it.

1.06. Exploitation of partial knowledge

When absolutely nothing can be assumed in advance about the behavior of the system, all an investigator can do is take random measurements and hope for the best. Fortunately, natural systems rarely behave so mysteriously, and usually a few things can be assumed before making any experiments. Thus the criterion of effectiveness is often a continuous function of the independent variables; many times it can be assumed to have but one peak in the region of interest. This book shows how to exploit such partial knowledge to develop efficient search techniques. In order to do this we shall have to perform a bit of what the mathematicians call *analysis* of functions. This means characterizing, in precise, quantitative terms, such intuitive concepts as unimodality (single-peakedness), smoothness, convexity, and similarity. We shall also need to describe—and extend to many dimensions—such geometric or even geographic ideas as peak, valley, ridge, pass or saddle, curvature, and slope. Like any good explorer, we can find our goal sooner if we know something of the lay of the land.

1.07. Multimodality, constraint, and time

After scanning the table of contents the reader will perhaps notice that several interesting topics related to experimental optimization are not discussed. For example there is no attempt to tackle multimodal problems having more than one peak because such situations have not been studied with any success so far. Anyone confronted with such a problem must at present try to isolate the various peaks and explore each of them individually. Hopefully, methods for handling multimodality will be available soon.

Likewise, there is no discussion of constrained optimization problems, in which certain combinations of variables are forbidden which would otherwise be optimal. In spite of its importance, this topic was omitted because there are already many books and articles discussing it. Readers interested in constrained optimization problems would do well to examine the literature on *linear programming, nonlinear* or *mathematical programming,* and *dynamic programming,* some of which is listed in the bibliography at the end of this chapter.

We will confine ourselves here to static problems, those in which the system does not change as time passes. A great deal of research on the dynamic case, where changes do take place, is now under way; some of this is listed in the bibliography. Dynamic problems, often involving *adaptive control,* have not been discussed here because, although much progress has been made on them, there is still much to be done. Moreover, the theoretical background needed, being still too strong for engineering seniors, would be inappropriate here.

1.08. Representation and scaling

Before beginning a search one should devote a little time to the choice of a good representation of the function and to the choice of scales of measurement. We are paraphrasing here the very sensible remarks on these matters expressed by Buehler, Shah, and Kempthorne.† They distinguish between the *relationship* between variables and the various *representations* of the relationship. For example, if we are investigating the dependence of chemical process yield on the adjustable operating variables, pressure and temperature, and if y is yield in grams, p is pressure in

† R. J. Buehler, B. V. Shah, and O. Kempthorne, "Some Properties of Steepest Ascent and Related Procedures for Finding Optimum Conditions," Iowa State University Statistical Laboratory (April 1961), pp. 8–10, 18.

atmospheres, π is the natural logarithm of p, and t is absolute temperature in degrees Kelvin, then the following expressions are two different *representations* of the same *relationship.*

$$y = \varphi(p, t) = y_0[1 - a(p - p_0)^2 - b(t - t_0)^2] \qquad (1\text{-}1a)$$

$$y = \psi(\pi, t) = y_0[1 - a(e^\pi - p_0)^2 - b(t - t_0)^2] \qquad (1\text{-}1b)$$

A good rule to follow is to choose a representation that can be approximated readily, at least in the neighborhood of the optimum, by a fairly low degree Taylor expansion. This is because most search techniques involve constructing approximations from measured estimates of first and second derivatives. By this rule, the quadratic representation Eq. (1-1a) is preferable to the other one involving the transcendental e^π term.

Another good rule is to prefer representations in which the factors do not interact. This criterion is satisfied by both Eqs. (1-1), since there is no term involving products of the two factors. However, the following representation, in which

$$x_1 = \tfrac{1}{2}(p - p_0) + \tfrac{1}{2}(t - t_0)$$

and

$$x_2 = \tfrac{1}{2}(p - p_0) - \tfrac{1}{2}(t - t_0)$$

shows interaction between x_1 and x_2 because of the term x_1x_2.

$$y = \theta(x_1, x_2) = y_0[1 - (a + b)(x_1^2 + x_2^2) - 2(a - b)x_1x_2] \qquad (1\text{-}2)$$

Methods for removing interaction are described in Secs. 3.16 and 5.16.

Finally it is good to select scales of measurement in which a unit change in one factor *at the optimum* gives the same change in the dependent variable as a unit change in any other factor. Thus if representation of Eq. (1-1a) were used we would want the scales of p and t to be such that

$$a(p - p_0)^2 = b(t - t_0)^2 \qquad (1\text{-}3)$$

This scaling rule, developed for a particular search procedure described in Sec. 4.13, turns out to be suitable for other procedures as well.

In geometric terms, these rules together tend to make the contours of the dependent variable spherical—symmetrical in all factors. It would of course rarely be the case that enough is known about the function to permit application of the rules. They are intended merely as a guide to educated guesses that tend, if accurate, to speed up the search procedure. Figure 1-1 contrasts a good and a bad choice of scale and representation.

1.09. Plausible reasoning

Before proceeding further, permit us a comment about the manner of presentation. Since we would like to see optimization theory covered in

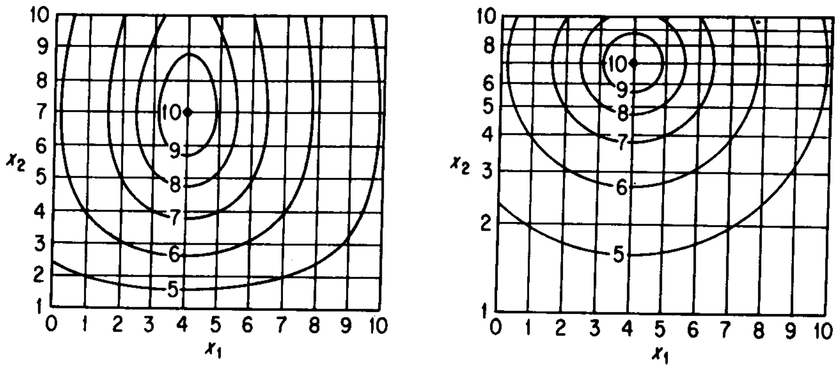

Fig. 1-1. (a) Original rectangular coordinates. (b) Preferable semi-logarithmic coordinates with scale change.

undergraduate engineering or economics curricula, where the students are not always used to formal logic, we have tried to make our definitions precise and our arguments rigorous. But, following the method of G. Polya† in his books on plausible reasoning, we have avoided the usual mathematical mode of exposition involving terse statement of formal theorems followed by formal proofs. As one of our purposes is to stimulate the student's creative imagination, we build up the results in the more or less intuitive manner of the original researcher groping toward new concepts. The reader who is more concerned with results than with how they are obtained will do well to skip over some of the details given. But the man who might want to do research in optimization theory—and there's much to be done—may prefer to develop his "feel" for the subject by closely following our plausible, if not always elegant, exposition and by working out formal proofs on his own.

SELECTED BIBLIOGRAPHY ON RELATED SUBJECTS

STEADY STATE CONSTRAINED OPTIMIZATION

Aris, R., *The Optimal Design of Chemical Reactors*. New York: Academic Press, 1961.

Hadley, G., *Linear Programming*. Reading, Mass.: Addison-Wesley Publishing Co., Inc., 1962.

Vajda, S., *Mathematical Programming*. Reading, Mass.: Addison-Wesley Publishing Co., Inc., 1961.

† G. Polya, *How to Solve It* (Princeton: Princeton University Press, 1945).
———, *Induction and Analogy in Mathematics* (Princeton: Princeton University Press, 1954).